

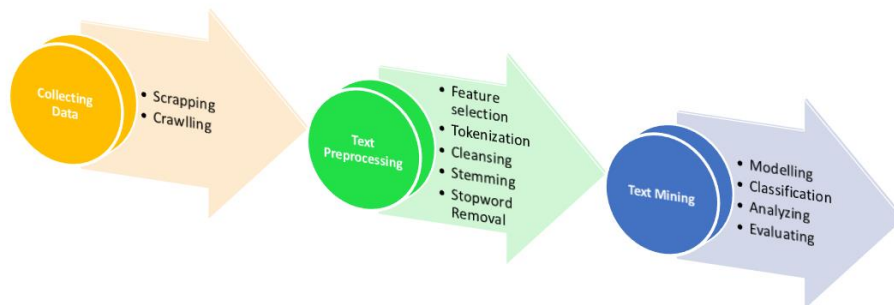
BAB II

LANDASAN TEORI

2.1. *Sentiment Analysis*

Sentiment analysis merupakan suatu metode atau teknik dalam melakukan *opinion mining* atau mengekstrak informasi untuk melihat pandangan seseorang terhadap suatu objek, apakah itu pandangan yang bersifat positif, negatif, atau netral. Analisis ini dilakukan untuk mengetahui bagaimana opini publik terhadap suatu kejadian, peristiwa, serta objek yang menghasilkan tingkat kepuasan. Analisis sentimen bertujuan untuk menganalisis isi atau makna dari teks [5].

Pendapat mengenai suatu produk oleh publik, khususnya pengguna sosial media menjadi tolak ukur untuk pengambilan keputusan bisnis di era teknologi. Penggunaan analisis sentimen sudah banyak digunakan diberbagai sektor bisnis seperti bidang perfilman, *online retail*, kosmetik, *fashion*, dan lainnya. Perusahaan memanfaatkan *feedback* dari users terhadap layanan atau produk untuk dijadikan strategi pemasaran dalam pengambilan keputusan.



Gambar 2.1. Langkah-langkah Sentiment Analysis [6]

Gambar 2.1 menunjukkan langkah-langkah secara umum analisis sentimen yang dibuat dengan mengacu kepada penelitian terdahulu dan telah dimodifikasi [6]. Berikut penjelasan dari tiap-tiap proses yang dilakukan:

1. Collecting Data

Tahap awal dalam melakukan analisis sentimen adalah mengumpulkan data dari sumber data yang telah ditentukan (sosial media, website, dll). Metode yang dapat digunakan antara lain adalah dengan menggunakan *web crawler* atau *web scraper* untuk menarik data dari sosial media (twitter, instagram, dll), tergantung seberapa besar data yang ingin ditarik.

2. *Text Preprocessing*

Setelah data terkumpul, data tersebut akan diolah sebelum dimasukkan ke dalam model untuk dilakukan analisis sentimen. Tahap ini merupakan tahap pembersihan data dari segala aspek yang tidak diperlukan, sehingga data siap untuk digunakan.

3. *Text Mining*

Langkah ini berisikan tahap pembuatan model dan tahap evaluasi. Data yang sudah siap untuk digunakan akan dimasukkan ke dalam model untuk mendapatkan informasi dari data. Tujuan dari tahap ini adalah untuk membangun sebuah model yang dapat memberikan penilaian opini terhadap data teks dan melihat performa dari model tersebut.

2.2. *Influencer Marketing*

Influencer marketing adalah strategi dari *digital marketing* dalam menarik *customers*, dengan memanfaatkan bantuan dari jasa *influencer* untuk memasarkan atau mempromosikan produk bisnis kepada para pengikut mereka. *Influencer* memiliki dampak yang besar dalam meningkatkan kepercayaan pelanggan terhadap suatu produk atau merk. Selain memperluas pemasaran produk, *influencer* juga meningkatkan *brand awareness* sehingga dapat memenangkan hati serta memberikan opini yang baik bagi para calon pelanggan [7].

Semakin berpesatnya perkembangan dan penggunaan teknologi, menyebabkan strategi *influencer marketing* menjadi elemen yang sangat penting bagi industri yang bergerak didunia digital. Kenaikan *influencer marketing* dari

tahun 2015 sampai tahun 2019 meningkat sebanyak 289%. *Influencer* dapat menghasilkan keuntungan bagi bisnis kurang lebih \$200 setiap kali melakukan *endorsement* sebagaimana kemampuan mereka dalam menarik *customers* [8].

2.3. Metode Naive Bayes

Naive Bayes merupakan algoritma *machine learning* yang dikemukakan oleh Thomas Bayes yang digunakan untuk klasifikasi yang menggunakan probabilitas *score based* dengan kondisi yang telah ditentukan. Algoritma ini belajar dari *training data* untuk memprediksi distribusi dari tiap kategori dan memberikan nilai probabilitas dengan *Bayes rule* [9].

$$P(A/B) = P(B/A) P(A) / P(B)$$

Rumus 2.1. Bayes Rule

Dengan metode tersebut, Naive Bayes dapat memberikan label bagi tiap-tiap *class* dengan persamaan:

$$label(d) = \operatorname{argmax} P(C_j)P(d | C_j)$$

Rumus 2.2. Naive Bayes Labelling

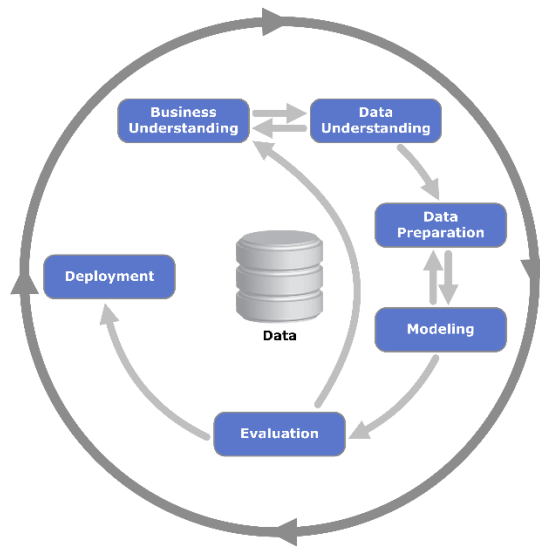
Beberapa penelitian telah dilakukan untuk membandingkan beberapa algoritma dalam melakukan klasifikasi teks dan Naive Bayes menjadi algoritma yang terbaik dengan menghasilkan akurasi terbaik [5]. Performance yang diberikan

oleh Naive Bayes sangat baik dalam melakukan klasifikasi dengan melihat keterikatan antar kata yang ada di dalam sebuah dokumen. Naive Bayes dapat menyelesaikan masalah dalam *multidimensional classification* yang dihadapi beberapa algoritma lainnya.

Model klasifikasi Naive Bayes juga tergolong mudah untuk diperbaharui secara bertahap karena pengimplementasiannya yang tergolong cukup sederhana. Ketika dilakukan pelatihan terhadap data yang ada dan kemudian ditambahkan data yang baru, maka fitur statistik yang ada akan diperbarui dan evaluasi tambahan akan dilakukan tanpa melakukan pengulangan terhadap *data training* yang lama. Naive Bayes juga tidak memerlukan proses generalisasi atau formulasi yang kompleks, tidak seperti algoritma lain [10].

2.4. CRISP-DM (*Cross-Industry Standard Process for Data Mining*)

Data mining telah menjadi sebuah teknik yang penting di dalam sebuah penelitian dengan kemampuan menganalisis data, menentukan relasi dan *pattern*, serta menemukan *insight* pada segudang data yang disimpan dalam *database* ataupun *warehouse*. Salah satu metode *processing* dalam melakukan *data mining* adalah dengan model CRISP-DM. CRISP-DM merupakan singkatan dari *Cross-Industry Standard Process for Data Mining* merupakan sebuah model atau metode yang menyediakan proses standard untuk *data mining* demi menyelesaikan permasalahan bisnis. Setiap tahap yang dilakukan di dalam CRISP-DM tergolong mudah digunakan dengan struktur tahapan yang jelas dan dokumen yang lengkap dan mendukung [11].



Gambar 2.2. Cross-Industry Standard Process for Data Mining [12]

Gambar 2.2 menjelaskan bahwa di dalam CRISP-DM terdapat 6 tahapan untuk memahami dan mengolah data serta beberapa tahap pengulangan di dalamnya sesuai dengan kebutuhan developer [13].

1. *Business Understanding*

Pada tahap ini dilakukan pemahaman terkait objektif yang ingin dicapai pada proses *data mining* atau penelitian yang dilakukan. Hal ini bertujuan untuk memberikan wawasan kepada developer mengenai apa yang selanjutnya harus dilakukan agar objektif tersebut dapat dicapai dengan maksimal. Tahap ini menyangkut pertukaran pikiran, pembacaan sumber referensi atau dokumen terkait, serta beberapa rumusan masalah terkait penelitian.

2. *Data Understanding*

Objektif yang telah digunakan pada tahap *business understanding* akan menjadi tolak ukur pengambilan data yang dibutuhkan dalam penelitian. Pada tahap *data understanding* ini dilakukan pemilihan data yang akan digunakan serta penarikan dataset yang dibutuhkan dari sumber-sumber terkait. Tahap ini juga menyangkut pemahaman terkait data baik dari segi kualitas, kebenaran serta keaslian, distribusi, dan tiap-tiap *fields* atau kolom yang ada. Data yang telah dikumpulkan nantinya akan diproses atau diolah untuk menjadi alat uji model yang akan digunakan.

Tahap ini merupakan tahap yang krusial karena keseluruhan data akan sangat mempengaruhi output yang akan dihasilkan. Menurut Terdapat beberapa tugas di dalam tahapan ini antara lain *collecting initial data*, *describe data*, *explore data*, *verify data quality* [14].

3. *Data Preparation*

Pada tahap ini, dataset yang telah dikumpulkan akan dilakukan pengolahan dengan tujuan menghasilkan data yang baik untuk digunakan dalam perhitungan atau algoritma. Langkah-langkah yang dilakukan pada tahap ini seperti menghilangkan *missing values*, menghilangkan data yang redundansi, standarisasi, serta pemilihan *fields* yang akan digunakan. Pada tahap ini bukan hanya dilakukan *cleansing* data tapi juga dilakukan *reformatting data* dan tahap integrasi data apabila terdapat lebih dari satu dataset yang digunakan [11].

4. *Modelling*

Pada tahap *modelling* akan ditentukan algoritma apa yang ingin digunakan untuk mencapai objektif berdasarkan parameter yang ditentukan, seperti contohnya *neural network*, *regression*, dan sebagainya. Model digunakan dengan menggunakan data yang telah di pisahkan menjadi *data training* dan *data testing*. *Data training* berperan sebagai alat perhitungan pembangunan model dengan algoritma yang telah ditentukan, sedangkan *data testing* berperan sebagai alat uji coba model yang telah digunakan untuk dilihat performanya. Pada tahap ini juga dapat dilakukan perbandingan antara beberapa model untuk dilihat performa terbaik yang dihasilkan untuk mencapai objektif yang telah ditentukan. Pada penelitian Naive Bayes dipilih sebagai algoritma yang digunakan untuk membangun model.

5. *Evaluation*

Pada tahap ini, model yang sudah digunakan akan dilakukan evaluasi untuk melihat apakah model tersebut sudah cukup baik atau masih kurang baik. Apabila model yang digunakan menghasilkan tingkat evaluasi yang kurang baik, maka tahap berikut dapat kembali ke tahap *modelling* atau *data preparation* untuk melakukan pengulangan tahap *data preprocessing*. Untuk mengevaluasi performa model *classifier*, terdapat beberapa teknik yang dapat digunakan antara lain:

a. *Accuracy*

Akurasi merupakan persentase tingkat kebenaran model dalam melakukan klasifikasi terhadap data uji (*testing*).

$$Accuracy = \frac{\text{jumlah prediksi benar}}{\text{jumlah total terprediksi}}$$

Rumus 2.3. Accuracy

Menurut rumus 2.3, secara matematika akurasi dapat dihitung sebagai jumlah prediksi benar dibagi dengan jumlah total terprediksi yang nantinya akan menghasilkan *rate* dari ketepatan klasifikasi [15].

b. *Confussion matrix*

Confussion matrix merupakan sebuah tabel yang berisikan *summary* dari prediksi atau klasifikasi yang dilakukan oleh model. Pada tabel tersebut dibagi ke dalam 4 kategori [16] yaitu:

- *True Positive (TP)*

Prediksi kelas positif yang benar

- *True Negative (TN)*

Prediksi kelas negatif yang benar

- *False Positive (FP)*

Prediksi kelas positif yang salah

- *False Negative (FN)*

Prediksi kelas negatif yang salah

Hasil dari *summary* tersebut divisualisasikan ke dalam bentuk tabel untuk dianalisa terkait benar atau tidak benarnya model dalam melakukan klasifikasi atau prediksi. Dengan mengetahui nilai dari kelas prediksi, maka juga didapat nilai *precision* dan *recall*. *Precision* merupakan persentase nilai *True Positive* dibandingkan dengan seluruh nilai prediksi positif [17].

$$Precision = \frac{TP}{TP+FP} * 100\%$$

Rumus 2.4. Precision

Sedangkan *recall* atau sensitifitas merupakan nilai dari prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Jika *precision* meningkat, maka hal tersebut akan mempengaruhi dalam penurunan nilai *recall*.

$$Recall = \frac{TP}{TP+FN} * 100\%$$

Rumus 2.5. Recall

Nilai *precision* dan *recall* dapat dirata-ratakan dengan menggunakan *harmonic mean*, sehingga menghasilkan nilai *F1-Score*.

$$F1-Score = 2 * \frac{precision*recall}{precision+recall}$$

Rumus 2.6. F1-Score

6. *Deployment*

Apabila model yang digunakan telah memnuhi kriteria sebagai model yang baik, maka model tersebut akan dilanjutkan ke tahap *deployment*. Pada tahap ini akan dibuat sebuah *engine* yang memuat model tersebut untuk dapat digunakan oleh *users*. Tahap ini menyangkut proses *planning* untuk mendeploy, memonitor dan melakukan maintenance, serta membuat report [14].

2.5. *Tweets*

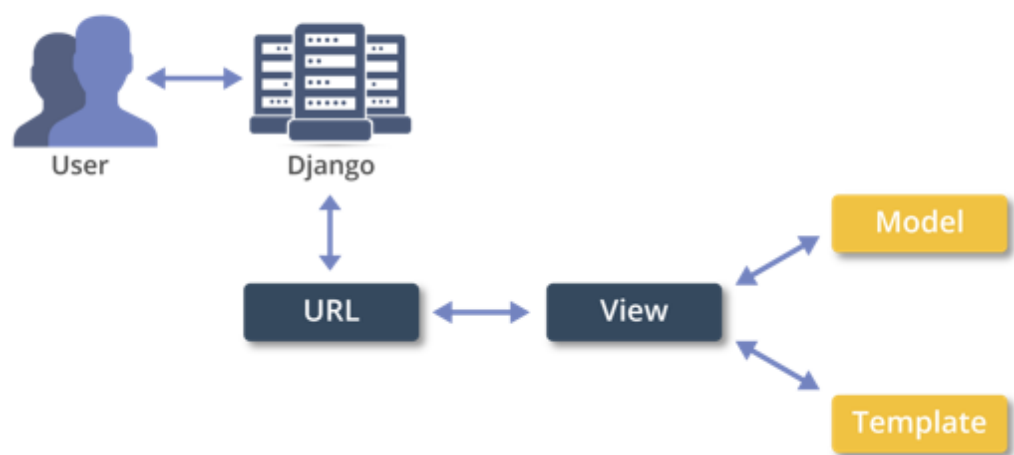
Tweets atau kicauan adalah unggahan yang dilampirkan oleh pengguna Twitter berupa teks sebanyak maksimal 280 karakter, sebagai bentuk pengekspresian diri dalam bentuk opini, komentar, maupun kritik terhadap sesuatu yang sedang terjadi maupun yang akan terjadi. Bahkan dilansir dari Wikipedia, Twitter tidak hanya berisikan pesan pengguna melainkan banyak tautan yang menarik, renungan mengenai masa depan dan pendidikan, rekomendasi musik, dan lain sebagainya [18]. Dalam melakukan penelitian *sentiment analysis* ini, *tweets* merupakan objek utama yang akan diteliti untuk mengambil makna serta opini yang diunggah oleh masyarakat Twitter.

2.6. *Django Framework*

Django merupakan suatu *framework* dengan bahasa pemrograman Python yang digunakan untuk membuat aplikasi berbasis web. Django memudahkan *full-stack developer* dalam menangani *front-end* dan juga *back-end* tanpa harus menuliskan kode dari nol [19]. Awalnya *framework* ini dikembangkan pada tahun

2003 dan 2005 oleh beberapa *developer* yang membangun web portal, dengan membangun *design pattern* pengembangan website secara umum.

Model View Template



Gambar 2.3. Konsep Django MVC [20]

Gambar 2.2 menunjukkan bahwa Django menerapkan konsep kerja MVT (*Model View Template*) dalam membantu menghubungkan sisi *front-end* dengan *back-end*. MVT memiliki cara kerja yang sama dengan model MVC (*Model View Controller*), dimana MVT merupakan modifikasi dari MVC.

2.7. Penelitian Terdahulu

Tabel 2.1. Penelitian Terdahulu

No	Penulis	Judul	Nama Jurnal	Hasil/Kesimpulan
1	Wongkar & Angdresey, 2019	<i>Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler : Twitter</i>	<i>Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019</i>	Metode Naive Bayes menghasilkan hasil yang terbaik yaitu akurasi sebesar 80,90%, dibandingkan dengan beberapa metode lain seperti KNN (75,58%) dan SVM (63,99%).

No	Penulis	Judul	Nama Jurnal	Hasil/Kesimpulan
2	Windasari, Ike Pertiwi Uzzi, Fajar Nurul Satoto, Kodrat Iman, 2017	<i>Sentiment Analysis on Twitter Posts: An analysis of Positive or Negative Opinion on GoJek</i>	<i>Proceedings - 2017 4th International Conference on Information Technology, Computer, and Electrical Engineering, ICITACEE 2017</i>	Penelitian ini menghasilkan akurasi sebesar 86% dengan <i>error</i> <i>rate</i> 14%. Tingkat prediksi benar positif adalah 100% dan tingkat prediksi benar negatif sebesar 67,44%, dengan menggunakan metode Naive Bayes.

No	Penulis	Judul	Nama Jurnal	Hasil/Kesimpulan
3	Gunawan, Billy Pratiwi, Helen Sasty Pratama, Enda Esyudha, 2018	Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes	JEPIN (Jurnal Edukasi dan Penelitian Informatika), Vol. 4, No. 2, Desember 2018	Analisis sentimen terhadap ulasan produk <i>online</i> dengan metode Naive Bayes menghasilkan akurasi 52,66% dengan jumlah <i>training data</i> 80% dan pengujian 5 <i>features</i> . Sedangkan dengan <i>training data</i> sebesar 90% dan pengujian 3 <i>features</i> , model yang dibuat menghasilkan akurasi sebesar 77,78%

No	Penulis	Judul	Nama Jurnal	Hasil/Kesimpulan
4.	Putra, Rein Rachman Johan, Monika Evelin Kaburuan, Emil Robert, 2019	<i>A Naïve Bayes Sentiment Analysis For Fintech Mobile Application User Review In Indonesia</i>	<i>International Journal of Advanced Trends in Computer Science and Engineering</i>	Penelitian ini melakukan komparasi performa model Naive Bayes dengan 2 dataset berbeda terkait <i>Fintech Application user review</i> yang dilakukan proses <i>cleansing</i> dan yang tidak. Penelitian ini menghasilkan model dengan akurasi 78% untuk dataset A dan 74% untuk dataset B dengan proses <i>cleansing</i> .

Berdasarkan penelitian-penelitian terdahulu yang dilampirkan pada tabel 2.1, maka penelitian untuk melakukan analisis sentimen terhadap produk yang

diiklankan oleh *influencer* akan dilaksanakan. Dengan adanya penelitian-penelitian terdahulu, maka diputuskan untuk menggunakan Naive Bayes sebagai metode penelitian *sentiment analysis* ini untuk melakukan pengujian serta evaluasi terhadap hasil yang dikeluarkan oleh model pada kasus ini, seperti yang dilakukan penelitian nomor 1 dan 3 pada tabel 2.1. Hal ini diputuskan melihat dari perbandingan beberapa metode yang telah dilakukan oleh penelitian terdahulu. Penelitian ini juga mengadopsi langkah yang dilakukan oleh penelitian terdahulu dengan melakukan *manual labelling* serta membagi *training* data dan *testing* data sebesar 90% dan 10% setelah dilakukan *cleansing data* seperti dalam penelitian nomor 3 dan 4 pada tabel 2.1.